

General H -theorem and entropies that violate the second law

Alexander N. Gorban

Department of Mathematics, University of Leicester, United Kingdom

We study systems with finite number of states A_i , which obey the first order kinetics (master equation). A general criterion is found for the existence of H -theorem with given H . A convex function H is a Lyapunov function for all master equations with the given equilibrium if and only if its conditional minima properly describe the equilibria of pair transitions $A_i \rightleftharpoons A_j$. This theorem does not depend on the principle of detailed balance and is valid both for reversible and for general Markov kinetics. Analysis of pair equilibria demonstrates, for example, that the popular Bregman divergences like Euclidian distance or Itakura–Saito divergence in the space of distributions cannot be the universal Lyapunov functions for the first-order kinetics and increase in some Markov processes.

The first non-classical entropy was proposed by Rényi in 1960 [1]. In the same paper, he discovered the very general class of divergences, the so-called f -divergences (or Csiszár–Morimoto divergences because the works of [2, 3] published simultaneously in 1963):

$$H_h(p) = H_h(P||P^*) = \sum_i p_i^* h\left(\frac{p_i}{p_i^*}\right) \quad (1)$$

where $P = (p_i)$ is a probability distribution, P^* is an equilibrium distribution, $h(x)$ is a convex function defined on the open ($x > 0$) or closed $x \geq 0$ semi-axis. We use here the notation $H_h(P||P^*)$ to stress the dependence of H_h both on p_i and p_i^* .

These divergences have the form of the relative entropy or, in the thermodynamic terminology, the (negative) free entropy, the Massieu–Planck functions [4], or F/RT where F is the free energy. They measure the deviation of the current distribution from the equilibrium.

After 1961, many new entropies and divergences were invented and applied to real problems, including Burg entropy [5], Cressie–Red family of power divergences [8], Tsallis entropy [6, 7], families of α -, β - and γ -divergences [9] and many others (see the review papers [10, 11]). Many of them have the f -divergence form, but some of them have not. For example, the squared Euclidean distance from P to P^* is not, in general, a f -divergence unless all p_i^* are equal (equidistribution). Another example gives the Itakura–Saito divergence:

$$\sum_i \left(\frac{p_i}{p_i^*} - \ln \frac{p_i}{p_i^*} - 1 \right). \quad (2)$$

The idea of Bregman divergences [12] provides a new general source of divergences that may differ from the f -divergences. Any strictly convex function F in a closed convex set V satisfies the Jensen inequality

$$D_F(p, q) = F(p) - F(q) - (\nabla_q F(q), p - q) > 0,$$

if $p \neq q$, $p, q \in V$. The positive quantity $D_F(p, q)$ is the Bregman divergence associated with F . For example, for a positive quadratic form $F(x)$ the Bregman di-

vergence is just $D_F(p, q) = F(p - q)$. In particular, if F is the squared Euclidean length of x then $D_F(p, q)$ is the squared Euclidean distance. If F is the Burg entropy, $F(x) = -\sum_i \ln p_i$, then $D_F(p, q)$ is the Itakura–Saito divergence. The Bregman divergences have many attractive properties. For example, the mean vector minimizes the expected Bregman divergence from the random vector [13]. The Bregman divergences are convenient for numerical optimization because the generalized Pythagorean identity [14].

For information processing and for many physical applications one more property is crucially important. *The divergence between the current distribution and equilibrium should monotonically decrease in Markov processes.* This is the ultimate requirement for use of the divergence in information processing and in non-equilibrium thermodynamics and kinetics. In physics, the first result of this type was Boltzmann’s H -theorem proven for nonlinear kinetic equation. In information theory, Shannon [15] proved this theorem for the entropy (“the data processing lemma”) and Markov chains. The H -theorem corresponds to the second law of thermodynamics. The data processing lemma reflects the general principle: information does not increase in random manipulations.

In his well-known paper [1], Rényi also proved that $H_h(P||P^*)$ monotonically decreases in Markov processes (he gave the detailed proof for the classical relative entropy and then mentioned that for the f -divergences it is the same). This result, elaborated further by Csiszár [2] and Morimoto [3] embraces many later particular H -theorems for various entropies including the Tsallis entropy and the Rényi entropy (because it can be transformed into the form (1) by a monotonic function, see for example [11]). The generalized data processing lemma was proven [16, 17]: For every two positive probability distributions P, Q the divergence $H_h(P||Q)$ decreases under action of a stochastic matrix $A = (a_{ij})$

$$H_h(AP||AQ) \leq \bar{\alpha}(A) H_h(P||Q),$$

$$\text{where } \bar{\alpha}(A) = \frac{1}{2} \max_{i,k} \left\{ \sum_j |a_{ij} - a_{kj}| \right\}$$

is the ergodicity contraction coefficient, $0 \leq \bar{\alpha}(A) \leq 1$. Here, neither Q nor P must be the equilibrium distribution: divergence between any two distributions decreases in Markov processes.

Under some additional conditions, the property to decrease in Markov processes characterizes the f -divergencies [18, 19]. For example, if a divergence decreases in all Markov processes, does not change under permutation of states and can be represented as a sum over states (has the trace form) then it is the f -divergence [11, 18].

Dynamics of distributions in the continuous time Markov processes is described by the master equation. Thus, the f -divergencies are the Lyapunov functions for the master equation. The important property of the divergencies $H_h(P\|P^*)$ is that they are the *universal* Lyapunov functions. That is, they depend on the current distribution P and on the equilibrium P^* but do not depend on the transition probabilities directly. Of course, without additional conditions like the trace form, the class of the universal Lyapunov functions for the master equations is much wider. For each new divergence we have to analyze its behavior in Markov processes and to prove or refute the H -theorem. For this purpose, we need a general and constructive criterion. It is desirable to avoid any additional requirements like the trace form or symmetry. In this paper we develop this criterion.

Obviously, the equilibrium P^* is a global minimum of any Lyapunov function $H(P)$ in the simplex of distributions. In brief, the *general H -theorem* states that a convex function $H(P)$ is a universal Lyapunov function for the master equation if and only if its conditional minima correctly describe the partial equilibria for pairs of transitions $A_i \rightleftharpoons A_j$. These partial equilibria are given by proportions $p_i/p_i^* = p_j/p_j^*$. They should be solutions to the problem

$$\begin{aligned} H(P) \rightarrow \min \quad & \text{subject to } p_k \geq 0 \ (k = 1, \dots, n), \\ \sum_{k=1}^n p_k &= 1, \quad \text{and given values of } p_l \ (l \neq i, j). \end{aligned} \quad (3)$$

We prove this general H -theorem and then analyze several Bregman's divergencies, which are not the f -divergencies, and demonstrate that they do not allow H -theorem even for systems with 3 states.

Three forms of master equation and the decomposition theorem. We consider continuous time Markov chains with n states A_1, \dots, A_n . The *master equation* for the probability distribution $P = (p_i)$ is

$$\frac{dp_i}{dt} = \sum_{j, j \neq i} (q_{ij}p_j - q_{ji}p_i) \quad (i = 1, \dots, n), \quad (4)$$

where q_{ij} ($i, j = 1, \dots, n, i \neq j$) are non-negative. In this notation, q_{ij} is the *rate constant* for the transition $A_j \rightarrow A_i$. Any set of non-negative coefficients q_{ij} ($i \neq j$) corresponds to a master equation. Therefore, the class

of the master equations can be represented as a non-negative orthant in $\mathbb{R}^{n(n-1)}$ with coordinates q_{ij} ($i \neq j$). Equations of the same class describe any first order kinetics in perfect mixtures.

Now, let us restrict our consideration to the set of the Markov chains with the given positive equilibrium distribution P^* ($p_i^* > 0$).

$$\sum_{j, j \neq i} q_{ij}p_j^* = \left(\sum_{j, j \neq i} q_{ji} \right) p_i^* \quad \text{for all } i = 1, \dots, n. \quad (5)$$

We join the transitions $A_i \rightleftharpoons A_j$ in pairs (say, $i > j$) and introduce the *stoichiometric vectors* γ^{ji} with coordinates:

$$\gamma_k^{ji} = \begin{cases} -1 & \text{if } k = j, \\ 1 & \text{if } k = i, \\ 0 & \text{otherwise.} \end{cases} \quad (6)$$

Let us rewrite the master equation (4) in the *quasichemical form*:

$$\frac{dP}{dt} = \sum_{i>j} (w_{ij}^+ - w_{ij}^-) \gamma^{ji}. \quad (7)$$

where $w_{ij}^+ = q_{ij}p_j^* \frac{p_j}{p_j^*}$ is the rate of the transitions $A_j \rightarrow A_i$ and $w_{ij}^- = q_{ji}p_i^* \frac{p_i}{p_i^*}$ is the rate of the reverse process $A_j \leftarrow A_i$ ($i > j$).

Systems with detailed balance form an important class of first order kinetics. The *detailed balance* condition reads: at equilibrium, $w_{ij}^+ = w_{ij}^-$, i.e.

$$q_{ij}p_j^* = q_{ji}p_i^* (= w_{ij}^*) \quad i, j = 1, \dots, n. \quad (8)$$

Here, w_{ij}^* is the *equilibrium flux* from A_i to A_j and back.

For the systems with detailed balance the quasichemical form of the master equation is especially simple:

$$\frac{dP}{dt} = \sum_{i>j} w_{ij}^* \left(\frac{p_j}{p_j^*} - \frac{p_i}{p_i^*} \right) \gamma^{ji}. \quad (9)$$

It is important that any set of non-negative equilibrium fluxes w_{ij}^* ($i > j$) defines a system with detailed balance (9) with a given positive equilibrium P^* . Therefore, the set of all systems (9) for any given equilibrium may be represented as a non-negative orthant in $R^{\frac{n(n-1)}{2}}$ with coordinates w_{ij}^* ($i > j$).

The *decomposition theorem* [20, 21] states that for any given positive equilibrium P^* and any positive distribution P the set of possible values dP/dt for equations (7) under the balance condition (5) coincides with the set of possible values dP/dt for equation (9) under the detailed balance condition (8). In other words, for every general system (7) with positive equilibrium P^* and any given non-equilibrium distribution P there exists a system with detailed balance (9) with the same equilibrium

and the same value of the velocity vector dP/dt at point P . Therefore, the sets of the universal Lyapunov function for the general master equations and for the master equations with detailed balance coincide.

General H -theorem. Let $H(P)$ be a convex function on the space of distributions. It is a Lyapunov function a master equations with the positive equilibrium P^* if $dH(P(t))/dt \leq 0$ for any positive solution $P(t)$. For a system with detailed balance (9)

$$\frac{dH(P(t))}{dt} = - \sum_{i>j} w_{ij}^* \left(\frac{p_j}{p_j^*} - \frac{p_i}{p_i^*} \right) \left(\frac{\partial H(P)}{\partial p_j} - \frac{\partial H(P)}{\partial p_i} \right) \quad (10)$$

The inequality $dH(P(t))/dt \leq 0$ is true for all non-negative values of w_{ij}^* , if and only if it holds for any term in (10) separately. That is, for any pair i, j ($i > j$) the convex function $H(P)$ is a Lyapunov function for the system (9) where only one w_{ij}^* is not zero.

A convex function on a straight line is a Lyapunov function for a 1D system with single equilibrium if and only if the equilibrium is a minimizer of this function. This elementary fact together with the previous observation gives us the criterion for universal Lyapunov functions for systems with detailed balance. Let us introduce the partial equilibria criterion:

Definition 1 A convex function $H(P)$ on the simplex of probability distributions satisfies the partial equilibria criterion with a positive equilibrium P^* if the proportion $p_i/p_i^* = p_j/p_j^*$ give the minimizers in the problem (3).

Remark 1 The partial equilibria criterion means that the partial equilibrium proportions $p_i/p_i^* = p_j/p_j^*$ give the minimizers in the problem (3) but it does not imply that these proportion give all the minimizers. If $H(P)$ is a convex but not strictly convex function then the minimizers for the given values of p_l ($l \neq i, j$) may form a segment which contains the partial equilibrium. Such a segment is a face of the level set of $H(P)$ in the plane with the given values of p_l ($l \neq i, j$). Good example give the divergencies

$$\begin{aligned} H_\infty(P||P^*) &= \max_i \left\{ \frac{p_i}{p_i^*} \right\} - 1 \text{ and} \\ H_{-\infty}(P||P^*) &= \max_i \left\{ \frac{p_i^*}{p_i} \right\} - 1. \end{aligned} \quad (11)$$

and their convex combinations [11].

Proposition 1 A convex function $H(P)$ on the simplex of probability distributions is a Lyapunov function for all master equations with the given equilibrium P^* which obey the principle of detailed balance if and only if it satisfies the partial equilibria criterion with this equilibrium.

Combination of this Proposition with the decomposition theorem [20] gives the same criterion for general master equations without hypothesis about detailed balance

Proposition 2 A convex function $H(P)$ on the simplex of probability distributions is a Lyapunov function for all master equations with the given equilibrium P^* if and only if it satisfies the partial equilibria criterion with this equilibrium.

These two Propositions together form the general H -theorem.

Theorem 1 The partial equilibria criterion with a positive equilibrium P^* is a necessary condition for a convex function to be the universal Lyapunov function for all master equations with detailed balance and equilibrium P^* and a sufficient condition for this function to be the universal Lyapunov function for all master equations with equilibrium P^* .

Let us stress that here the partial equilibria criterion provides a necessary condition for systems with detailed balance (and, therefore, it is necessary for more general systems without detailed balance assumption) and a sufficient condition for the general systems (and, therefore, for the narrower class of systems with detailed balance).

Examples. The simplest Bregman divergence is the squared Euclidean distance between P and P^* , $\sum_i (p_i - p_i^*)^2$. The solution to the problem (3) is: $p_i - p_i^* = p_j - p_j^*$. Obviously, it differs from the proportion required by the partial equilibria criterion (Fig. 1a). For the Itakura-Saito divergence (2) the solution to the problem (3) is: $\frac{1}{p_i} - \frac{1}{p_i^*} = \frac{1}{p_j} - \frac{1}{p_j^*}$. It also differs from the proportion required (Fig. 1b).

If the single equilibrium in 1D system is not a minimizer of a convex function H then $dH/dt > 0$ on the interval between the equilibrium and minimizer of H (or minimizers if it is not unique). Therefore, if $H(P)$ does not satisfy the partial equilibria criterion then in the simplex of distributions there exists an area bordered by the partial equilibria surface for $A_i \rightleftharpoons A_j$ and by the minimizers for the problem (3), where for some master equations $dH/dt > 0$ (Fig. 1). In particular, in such an area $dH/dt > 0$ for the simple system with two transitions, $A_i \rightleftharpoons A_j$, and the same equilibrium.

Discussion. Many non-classical entropies are invented and applied to various problems in physics and data analysis. The problem of H -theorem appears in many areas of research and is important for practical applications. A good example give us the discussion of H -theorem with the classical and non-classical entropies for the lattice Boltzmann methods [22, 23], a field of science between kinetic theory and numerical analysis.

We suggest that if an entropy has no H -theorem (that is, it violates the second law and the data processing lemma) then there should be unprecedentedly strong reasons for its use. Without such strong reasons we cannot employ it. In this paper, the general criterion for the existence of H -theorem is proved. It has a simple and physically transparent form: the convex divergence (relative entropy) should properly describe the partial equilibria for all pair transitions $A_i \rightleftharpoons A_j$. It is straightforward

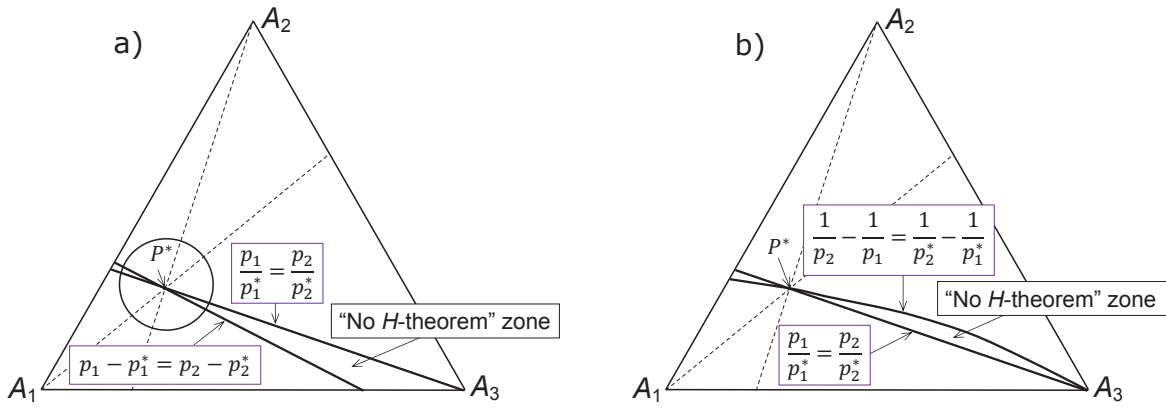


FIG. 1: The triangle of distributions for the system with three states A_1, A_2, A_3 and the equilibrium $p_1^* = \frac{4}{7}, p_2^* = \frac{2}{7}, p_3^* = \frac{1}{7}$. The lines of partial equilibria $A_i \rightleftharpoons A_j$ given by the proportions $p_i/p_i^* = p_j/p_j^*$ are shown, for $A_1 \rightleftharpoons A_2$ by solid straight lines (with one end at the vertex A_3), for $A_2 \rightleftharpoons A_3$ and for $A_1 \rightleftharpoons A_3$ by dashed lines. The lines of conditional minima of $H(P)$ (3) are presented for the partial equilibrium $A_1 \rightleftharpoons A_2$: (a) for the squared Euclidean distance this is a straight line given by the equation $p_1 - p_1^* = p_2 - p_2^*$ (a circle here is an example of the $H(P)$ level set), and (b) for the Itakura-Saito divergence this is a curve given by the equation $\frac{1}{p_2} - \frac{1}{p_1} = \frac{1}{p_2^*} - \frac{1}{p_1^*}$. Between these lines and the line of partial equilibria the “No H-theorem” zone is situated. In this zone, $H(P)$ increases in time for some master equations with equilibrium P^* . Similar zones (not shown) exist near other partial equilibrium lines too. Outside these zones, $H(P)$ monotonically decreases in time for any master equation with equilibrium P^* .

to check this partial equilibria criterion. The applicability of this criterion does not depend on the detailed balance condition and it is valid both for the class of the systems with detailed balance and for the general first order kinetics without this assumption. We describe the universal Lyapunov function for the master equations in

a form of the constructive criterion which allows us to test any convex function. It is possible to look for alternative description of the universal Lyapunov function in the form of constructive parametric representation. We can expect new promising classes of entropies from this direction of research in the future.

-
- [1] A. Rényi, in *Proceedings of the 4th Berkeley Symposium on Mathematics, Statistics and Probability 1960*; University of California Press: Berkeley, CA, USA, 1961; Vol. 1; 547–561.
 - [2] I. Csizsár, Magyar. Tud. Akad. Mat. Kutato Int. Kozl. **8**, 85–108 (1963).
 - [3] T. Morimoto, J. Phys. Soc. Jap. **12**, 328–331 (1963).
 - [4] H.B. Callen, *Thermodynamics and an Introduction to Thermostatistics* (Wiley, New York, 1985).
 - [5] J.P. Burg, Geophysics **37** **1972**, 375–376.
 - [6] C. Tsallis, J. Stat. Phys. **52**, 479–487 (1988).
 - [7] Abe, S., Okamoto, Y. Eds. *Nonextensive Statistical Mechanics and its Applications* (Springer, Heidelberg, 2001).
 - [8] N. Cressie, T. Read, J. R. Stat. Soc. Ser. B **46**, 440–464 (1984).
 - [9] A. Cichocki, S.-I. Amari, Entropy **12** (6) 1532–1568 (2010).
 - [10] M.D. Esteban, D. Morales, Kybernetika **31**, 337–346 (1995).
 - [11] A.N. Gorban, P.A. Gorban, G. Judge, Entropy **12** (5), 1145–1193 (2010). arXiv:1003.1377 [physics.data-an].
 - [12] L.M. Bregman, USSR Comput. Math. Math. Phys. **7** (3), 200217 (1967).
 - [13] A. Banerjee, S. Merugu, I.S. Dhillon, J. Ghosh, J. Machine Learning Res. **6**, 17051749 (2005).
 - [14] I. Csizsár, F. Matúš, 2012. arXiv:1202.0666 [math.OC]
 - [15] C.E. Shannon, Bell System Technical Journal **27**, 379–423 (1948).
 - [16] J.E. Cohen, Y. Derriennic, G.H. Zbaganu, Contemp. Math. **149**, 251–259 (1993).
 - [17] J.E. Cohen, Y. Iwasa, G. Rautu, M.B. Ruskai, E. Seneta, G. Zbaganu, Linear. Alg. Appl. **179**, 211–235 (1993).
 - [18] P. Gorban, Physica A **328**, 380–390 (2003). arXiv:cond-mat/0304131 [cond-mat.stat-mech].
 - [19] S.-I. Amari, in *Proceedings of the 16th International Conference on Neural Information Processing*; C.S. Leung, M. Lee, J.H. Chan, Eds., LNCS 5863, (Springer, Berlin, 2009); pp. 185–193.
 - [20] A.N. Gorban, Local equivalence of reversible and general Markov kinetics, Physica A, available online: 23-NOV-2012, DOI: 10.1016/j.physa.2012.11.028. arXiv:1205.2052 [physics.chem-ph].
 - [21] A.N. Gorban, Comput. Math. Appl., in press (2013). arXiv:1212.5142 [physics.data-an].
 - [22] I.V. Karlin, A.N. Gorban, S. Succi, and V. Boffi, Phys. Rev. Lett. **81**, 6–9 (1998).
 - [23] A.N. Gorban, D. Packwood, Phys. Rev. E **86**, 025701(R) (2012).